MIMIC: A Molecular-Field Matching Program. Exploiting Applicability of Molecular Similarity Approaches

JORDI MESTRES,* DOUGLAS C. ROHRER, GERALD M. MAGGIORA

Computer-Aided Drug Discovery, Pharmacia & Upjohn Inc., Kalamazoo, Michigan 49001

Received 19 July 1996; revised 2 October 1996

ABSTRACT: This contribution presents the development and applicability of MIMIC, a molecular-field matching program for quantitatively evaluating the similarity between molecules, in a computationally feasible way and assesses the relative orientation that maximizes their similarity. In the present version one can deal with two types of molecular-field similarities, namely steric volume and electrostatic, as well as a combined similarity that takes into account a given contribution from each of them. Besides optimization of the relative spatial orientation by maximizing these similarities, MIMIC can perform exhaustive searches to locate a global similarity maximum candidate and the set of local similarity maxima closest to it. The high accuracy of the approach permits evaluation of the similarity space coverage of each similarity maximum. In addition, the study of the relationships among similarity spaces is proposed as a strategy to better understand the linkage between the different molecular overlay solutions obtained from the use of different molecular-field representations. Finally, calculation of the atomic contributions to the total molecular similarity provides a means for locating maximum similarity loci and for constructing pharmacophore patterns. The methodological bases and a detailed description of how they were implemented in MIMIC to handle molecular matchings between large systems is presented. All current features of the program were applied to a relative ligand-binding problem between two nonnucleoside HIV-1 reverse transcriptase inhibitors. © 1997 by John Wiley & Sons, Inc. *J Comput Chem* **18**: 934–954, 1997

* *Permanent address:* Institut de Quimica Computacional, Universitat de Girona, 17071 Girona, Catalonia, Spain.

Correspondence to: D. C. Rohrer; e-mail: douglas.c.rohrer@am.pnu.com.

Contract grant sponsor: CIRIT (Generalitat de Catalunya; contract grant number 1995BEAI200115.

CCC 0192-8651 / 97 / 070934-21

Keywords: molecular similarity; molecular-field matching; conformational clustering; similarity space exploration; nonnucleoside HIV-1 reverse transcriptase inhibitors

Introduction

Molecular similarity is a difficult concept to define in a strict and general fashion. Consequently, over the last several years a large number of approaches have appeared that attempt to quantify this concept.[1–5] The relevance of the contribution of molecular similarity approaches during the process of designing a drug lead is still far from being fully exploited. However, it is obvious that they can become a powerful computational tool to predict relative ligand-binding orientations at the initial stages of drug design projects, when information about the binding pocket is not available or only bare indirect data are known.

Molecular overlay algorithms provide a means for determining superpositions between molecules that maximize some form of molecular-similarity function. This procedure is referred to as *molecular matching*. It is generally assumed that ligands and their molecular field characteristics will align in a similar way when binding to a receptor. The degree of complementarity between the molecular fields of a ligand and a receptor should be directly related to the binding strength and relative activity of the ligand. However, a question arises regarding which molecular fields, or which combination of molecular fields, are appropriate for describing the relevant structural and electronic features of a family of molecules to best reflect their biological activities.

Three types of molecular fields are generally used in 3-dimensional molecular field based similarity studies: *steric*, *electrostatic*, and *hydrophobic*. These fields attempt to provide information on different aspects of molecular similarity. The steric contribution reflects molecular size and overall shape, both features being important to the steric complementarity of a ligand at the binding site.[6–9] On the other hand, long-range interactions between a ligand and its binding site are likely to be electrostatic in nature. This leads to the hypothesis that electrostatic interactions may have, in some cases, an important contribution to the ligand/protein binding mechanism.[10–17] Finally, other electrostatic properties are the result of averaged interactions of the molecule with its surroundings (solvent and protein environment). This average is commonly represented by a phenomenological interaction such as the hydrophobicity (or lipophilicity). Different approaches have recently appeared in an effort to theoretically represent the hydrophobic interactions in terms of local solute–solvent electrostatics.[18–27]

The aim of this contribution is to present MIMIC, a new molecular field based matching program, and to illustrate some of its actual capabilities. Due to an increasing interest in this type of computational tool for molecular modeling and drug design, a variety of molecular-field matching programs have been developed over the last few years.[27–41] The main differences among them are the type of molecular fields they use and the ways they approach molecular field based similarities. Some of the programs also have the ability to improve the final similarity by allowing for conformational flexibility. This can be of importance when trying to determine the bioactive conformation, and several interesting examples have recently been reported.[41–45] Another important feature that characterizes molecular-matching programs is their ability to perform exhaustive searches of molecular similarity space to locate the relative orientations between two molecules having high similarity.[32] In fact, this turns out to be an important step in any matching process because it ultimately determines the final reliability of the results.

A detailed description of the methodological aspects upon which the program is based are presented first, followed by an example that describes the application of the methodology to a study of the relative binding orientation between two nonnucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) of considerably different structure, namely, Nevirapine and α-anilinophenylacetamide (α-APA).

Methodological Aspects

The main characteristics of MIMIC are presented in the following subsections. In particular, special attention has been paid to the description

of the molecular fields implemented, the molecular similarity definitions used, and other relevant molecular-matching utilities currently available in the program.

Molecular Fields

Two different molecular fields, which account for the steric and the electrostatic features of molecules, are used. Accurate calculations of these molecular fields can be performed and are routinely being done.[9,10] However, they are too computationally demanding for use during the matching of molecules in drug design. Thus, approximations to the description of these molecular fields are needed to speed-up molecular superposition calculations and enlarge the size and the number of the molecules where the molecular similarity approach can be applied. The approximations implemented in MIMIC are described in the following sections.

STERIC VOLUME FIELD

The molecular steric volume (MSV) field is defined according to a procedure previously described by Rohrer.[37] Under this approximation, the MSV of a molecule A is basically given by the simple additive model

$$F_A^{\mathrm{MSV}}(\mathbf{r}) = \sum_{i \in A} f_i^{\mathrm{MSV}}(\mathbf{r}), \tag{1}$$

where f_i^{MSV} is the atomic steric volume contribution of atom i of molecule A. Each f_i^{MSV} is represented by an atom-centered single-Gaussian function as

$$f_i^{\mathrm{MSV}}(\mathbf{r}) = \alpha_i \exp\left(-\beta_i |\mathbf{r} - \mathbf{R}_i|^2\right). \tag{2}$$

where $\mathbf{R}_i$ is the nuclear coordinate position of atom i, and the coefficient α_i and the exponent β_i are parameters that have to be optimized for each atom. Grant and Pickup[46] recently reported a similar approach using a Gaussian description of molecular shape to rapidly evaluate molecular areas and volumes. In addition to the ease of integration inherent with Gaussian functions, the use of this ''soft-Gaussian'' approach presents some relevant advantages during the molecular-matching process.

ELECTROSTATIC FIELD

To speed up calculations and be able to treat large molecules, the molecular electrostatic potential (MEP) is calculated using an atomic point-charge approximation. Under this approximation, the MEP of a molecule A at a point $\mathbf{r}$ is given as

$$F_A^{\mathrm{MEP}}(\mathbf{r}) = \sum_{i \in A} \frac{q_i}{|\mathbf{r} - \mathbf{R}_i|}, \tag{3}$$

where q_i and $\mathbf{R}_i$ are, respectively, the partial atomic charges and the nuclear coordinates of the constituent atoms of molecule A. Although this is a rather crude estimate, it is generally adequate to qualitatively reproduce most of the important electrostatic features of a molecule.[47]

Molecular Similarity

Once the molecular fields are defined, quantification of the similarity between the molecular fields of two or more molecules can be accomplished. A description of the similarity measures and indices currently available in MIMIC follows.

SIMILARITY MEASURES

One of the most widely used *molecular similarity measures* between two molecules was first reported by Carbó et al.[48] They defined a density-based quantum molecular similarity measure as

$$Z_{AB}(\mathbf{r}_A, \mathbf{r}_B; \theta) = \iint \rho_A(\mathbf{r}_1)\theta(\mathbf{r}_1, \mathbf{r}_2)\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2, \tag{4}$$

where ρ_A and ρ_B are, respectively, the electron density distributions of molecules A and B, $\theta(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite operator depending on two-electron coordinates, and $\mathbf{r}_A$ and $\mathbf{r}_B$ denote the coordinate sets of molecules A and B, respectively. If A and B are the same molecule, the value obtained from eq. (4) is called the molecular self-similarity measure.

The value of Z_{AB} depends upon two principal characteristics: the mutual positions and orientations of the two molecules being compared, and the nature of the positive definite operator θ. Thus, for a given θ, Z_{AB} can be maximized by translating and rotating B with respect to A until an optimum superposition of the two is obtained. On the other hand, as recently suggested,[49] use of different operators will lead to different contributions to the molecular similarity. For example,

when $\theta(\mathbf{r}_1, \mathbf{r}_2)$ is the Dirac delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$, eq. (4) is reduced to calculation of the overlap integral between the two density distributions giving what has been called overlap similarity. When $\theta(\mathbf{r}_1, \mathbf{r}_2)$ is substituted by $1/\mathbf{r}_{12}$ an electronic repulsion integral expression is obtained, yielding the so-called Coulomb similarity. It should be noted that what was originally formulated as a density-based quantum molecular similarity can be generalized by simply replacing the electron-density distributions in eq. (4) by any kind of molecular field. Thus, in MIMIC a molecular similarity measure is computed as

$$Z_{AB}(\mathbf{r}_A, \mathbf{r}_B) = \int F_A^{\mathrm{MF}}(\mathbf{r}) F_B^{\mathrm{MF}}(\mathbf{r})\, d\mathbf{r}, \qquad (5)$$

where MF denotes the type of molecular field (in the actual version MSV or MEP) from which Z_{AB} is obtained.

Here, calculation of Z_{AB} from eq. (5) using MEPs deserves some comment. The molecular electrostatic similarity cannot be computed directly from the electrostatic field given in eq. (3) because the MEP diverges at nuclear positions $\mathbf{R}_i$. This deficiency was removed by incorporating a nuclear-smoothing procedure, which eliminates divergencies at the nuclei. In terms of overall electrostatic similarity, the contribution from regions away from the nuclei play a much more important role than those at the nuclei due to the long-range attractive nature of the electrostatic interactions. Thus, nuclear smoothing only has a minor effect on the overall electrostatic similarity. In particular, MIMIC uses the smoothing approach proposed by Good et al.,[16] which replace the inverse distance-dependence term in eq. (3) by a sum of three Gaussian functions. Under this approximation, the molecular electrostatic similarity measure is computed as

$$\begin{aligned} Z_{AB}(\mathbf{r}_A, \mathbf{r}_B) = & \sum_{i \in A} \sum_{j \in B} q_i q_j \\ & \times \left[\int \left(\sum_{m=1,3} a_m \exp\left(-b_m |\mathbf{r} - \mathbf{R}_i|^2\right) \right) \right. \\ & \left. \times \left(\sum_{n=1,3} a_n \exp\left(-b_n |\mathbf{r} - \mathbf{R}_j|^2\right) \right) d\mathbf{r} \right], \end{aligned} \qquad (6)$$

where q and $\mathbf{R}$ are the atomic charges and the nuclear coordinates, respectively, and a and b are the coefficients and exponents of the three-Gaussian best fit to the $1/\mathbf{r}$ function.[16] Thus, calculation of Z_{AB} is reduced to the evaluation of nine simple integrals between two-center Gaussians.

Another interesting point to be considered is that, according to the definition of molecular similarity measure in eq. (5), the steric contribution of atom i of molecule A to the total MSV similarity measure between molecules A and B is given by[50]

$$Z_{iB}(\mathbf{r}_A, \mathbf{r}_B) = \int f_i^{\mathrm{MSV}}(\mathbf{r}) F_B^{\mathrm{MSV}}(\mathbf{r})\, d\mathbf{r}, \qquad (7)$$

where f_i^{MSV} and F_B^{MSV} are defined according to eqs. (1) and (2). Thus, it is possible, using these atomic contributions, to obtain quantitative information about where the structural similarity between two molecules is mainly concentrated. Location of such *maximum similarity loci* can then be used as a tool for constructing structural pharmacophore patterns.[51]

Expanding the possibilities of this methodology, a multiple MSV similarity measure among M molecules can be defined as[52]

$$Z_{\{M\}}(\mathbf{r}_{\{M\}}) = \int \prod_{I=1,M} F_I^{\mathrm{MSV}}(\mathbf{r})\, d\mathbf{r}. \qquad (8)$$

In MIMIC, a general code computes these $Z_{\{M\}}$, taking advantage of two mathematical tools: nested do loops,[52] which generally control the integrals to be computed, and a Gaussian representation of the molecular steric volume [eq. (2)], which permits application of the Gaussian-product theorem.[53] This makes the algorithm general, simple, and compact.

SIMILARITY INDICES

Once a molecular similarity measure has been defined, a *molecular similarity index* can be calculated. Although MIMIC can calculate different types of similarity indices[17,48,54–56] for comparative purposes, throughout this work the original form defined by Carbó et al.[48] is used. In this case, the pairwise molecular similarity index is evaluated as a generalized cosine of the angle between the two molecular field distributions,

$$S_{AB} = \frac{Z_{AB}}{(Z_{AA} \cdot Z_{BB})^{1/2}}. \qquad (9)$$

When S_{AB} is calculated from MSV-based similarity measures [using eqs. (1) and (5)], the positive definite nature of the MSV fields restricts the similarity index value between 0 (total dissimilarity)

and +1 (complete similarity or identity). However, when S_{AB} is calculated from MEP-based similarity measures [eq. (6)], the nonpositive definite nature of MEPs allows the similarity index to range from a value of −1 (total complementarity) to +1 (complete similarity). Note that eq. (9) can easily be generalized to calculate multiple MSV similarity indices using the multiple MSV similarity measures defined in eq. (8).

The molecular similarity indices calculated from eq. (9) belong to the class of what can be called *continuum indices*, because they come from an analytical integration of the overlap of molecular fields over all space. In addition to this definition, calculation of *cumulative indices* from molecular field values obtained from grid points embedding the ensemble of the molecules is also feasible and has been widely employed, especially in the evaluation of electrostatic similarities by avoiding nuclei discontinuities.[17] Under this approach, the grid-based Carbó molecular similarity index can be expressed as

$$S_{AB} = \frac{\sum_i F_A^i F_B^i}{\left(\sum_i F_A^i F_A^i\right)^{1/2}\left(\sum_i F_B^i F_B^i\right)^{1/2}}, \qquad (10)$$

where F symbolizes any value of the molecular field used to calculate the molecular similarity index and the summations extend over all grid points. Obviously, boundary conditions of cumulative indices depending upon the positive or nonpositive nature of the molecular field are the same as for the continuum indices mentioned earlier.

Two types of grid-based cumulative molecular similarity indices were implemented, depending on which molecular region accounted for the similarity. In the first approach, only those grid points outside the union of the van der Waals molecular volumes were considered when evaluating the similarity index. The interior region was considered empty. In this case, grid limits were defined to extend 5 Å away from the common Cartesian coordinate limits of the ensemble of the two molecules. The second approach accounted for all points in space, regardless of its situation inside or outside the union of the van der Waals molecular volumes. MSV fields are continuous, but MEPs present discontinuities at the nuclei and some distance cutoff around nuclei is necessary to avoid this problem. Thus, for cumulative electrostatic similarities, MIMIC uses the so-called *extended* MEP recently suggested by Rohrer.[37] In this case, instead of dealing with an empty MEP in regions inside the union of the van der Waals molecular volumes, this approach considers the atoms as hard spheres of radius 1 Å and constant atomic charge. Actually, this grid-based approximation turned out to be equivalent to the electrostatic similarity indices evaluated using the three-term Gaussian expansion [eq. (6)], because the three-term Gaussian fit to the inverse distance function begins to deviate around 1 Å.[16] Grid limits are now defined to extend 3 Å away from the common molecular Cartesian coordinate limits.

One advantage of the grid-based cumulative (when an adequate grid spacing is used) over continuum approaches arises when comparing multiple molecules. Computation of the grid-based cumulative multiple MSV similarity index scaled linearly with the number of molecules (for a given grid size) while the integral-based evaluation scaled exponentially [eq. (8)]. As a result, when the number of molecules being compared became larger, computing times of multiple molecule similarities could be reduced by using the grid-based approach without sacrificing accuracy.

Another approach to the molecular similarity index comes from considering a weighted combination of molecular similarities from different molecular fields,[32]

$$S_{AB} = \sum_{i \in F} w_i (S_{AB})_i, \quad \sum_{i \in F} w_i = 1. \qquad (11)$$

The summations in eq. (11) extend over the set of molecular fields (F) considered. In the current version of MIMIC this set includes the steric and electrostatic contributions. The w_i weighting parameters are user selected. In cases where structural experimental data for the complexes between different ligands and their common receptor are available, the weights corresponding to each molecular-field similarity contribution can be altered to reproduce the relative orientation of the ligands at the binding site. Weights obtained in this manner can then be used in cases where the ligand binding orientations are unknown.

Molecular Field Matchings

The molecular similarity matching procedure maximizes the corresponding molecular similarity measure or index from a given starting relative orientation of the molecules. The program is initially provided with the Cartesian coordinates of

the two molecules. Then the relative position and orientation are varied to optimize the match between the molecular fields using common steepest-descent or Newton–Raphson techniques. One molecule is kept fixed (the *reference molecule*) while the other is moved to optimize its orientation relative to the reference molecule (the *adapting molecule*).

The applicability of molecular-field matchings for drug design purposes has not yet been fully exploited. What follows are some of the utilities MIMIC can now handle. Other utilities are being developed and will be the subject of future studies.

CONFORMATIONAL CLUSTERING

Optimization of the molecular geometry using any of the current molecular mechanics force fields, or semiempirical, or *ab initio* quantum mechanical calculations in combination with conformational analysis are compulsory before facing a similarity study. In fact, the value of the similarity measure and, thus, the values of the corresponding similarity indices, depend on the 3-dimensional structure of molecules being compared [eq. (5)]. Consequently, the final similarity results are strongly affected by these molecular structures. For this reason, consideration of different conformations for each single molecule is being employed as a strategy for dealing with inherent structural flexibility.[57]

Dealing with the molecules of interest for drug design purposes sometimes presents the additional difficulty that, after performing the molecular conformational analysis, a vast number of conformations may have been generated. Obviously, a similarity study by considering each one of these conformations appears to be computationally unreasonable. Thus, to make similarity studies computationally feasible it would be useful to be able to select a subset of *prototype conformations* for each molecule for further consideration during the molecular-matching similarity study. Usually a clustering technique is employed to identify commonalities among the whole set of conformations and to reduce the number to a representative subset that covers the most significant regions of the conformational space.[58,59] In MIMIC this is accomplished by calculating all possible pairwise molecular similarities and constructing a molecular similarity matrix. Within this similarity matrix, each conformation is described as a vector in an M-dimensional space, where M is the number of conformations. The elements of each vector are the pairwise similarities. A principle component analysis is then used to reduce the data collected in the molecular similarity matrix to a smaller number of orthogonal variables that conserve the largest part of the initial similarity information. After clustering, the minimum energy conformation in each cluster is taken as a representative structure of the conformational space covered by the subset of conformations within the cluster.

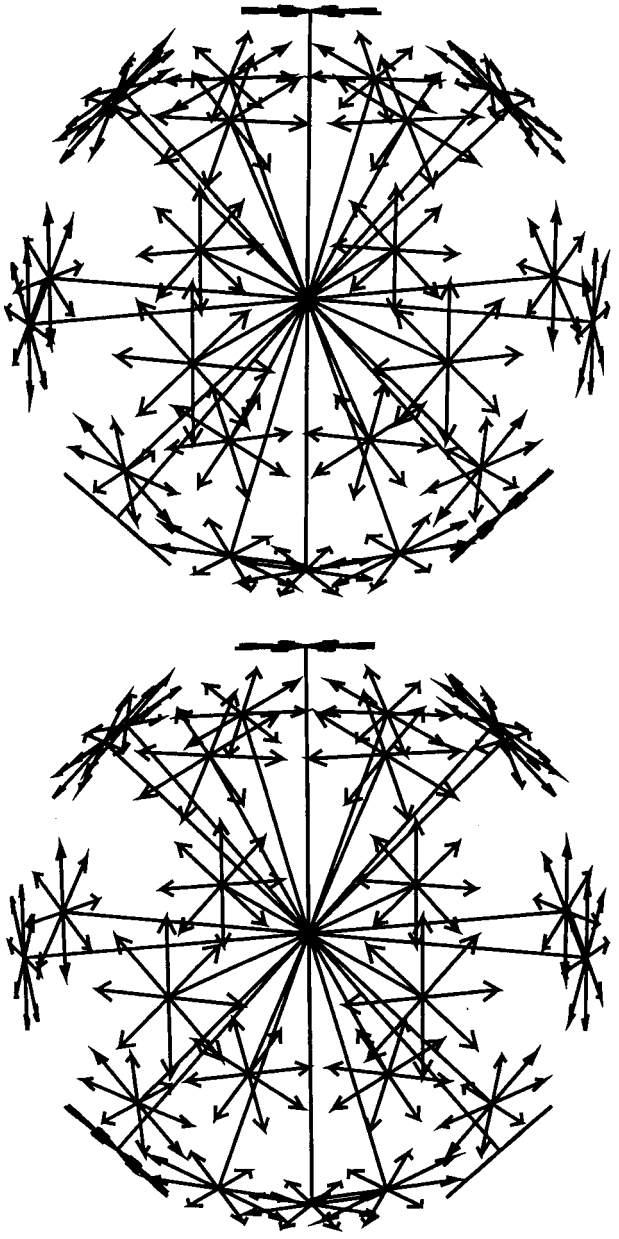
EXPLORATION OF SIMILARITY SPACE

Because a gradient-seeking technique is used in the optimization process, MIMIC will only attain the nearest similarity maximum from a given starting relative orientation. However, it is obvious that other starting points would likely coverage to different similarity maxima, depending upon the particular size and 3-dimensional atomic arrangement of the molecules.[32,60] Thus, one faces a typical multiple extrema problem.

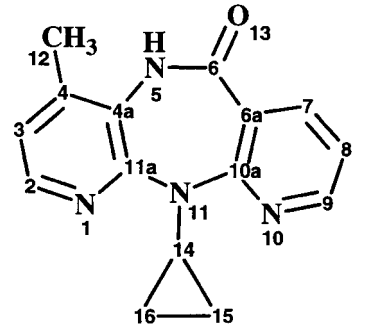
To tackle this problem, MIMIC uses a spherical systematic search algorithm to generate a large variety of initial relative orientations to locate the best similarity maximum and other similarity maxima close to it. Initially, the two molecules are translated so that their centers of mass are superimposed. Then, the adapting molecule is rotated to describe a complete sphere around the reference molecule to ensure a wide and uniformly distributed search of the molecular similarity space. This is schematically shown in Scheme 1, where the directional vectors are depicted using a rotational step of 45°, giving rise to 208 unique starting positions and orientations. The program also has the capability of describing several spheres of different radius to account for possible molecular size and shape features. The high accuracy of the methodology permits accumulation of the frequency that a given maximum is found from various starting points, which provides an idea of the similarity space coverage of each similarity maximum. This information, together with the study of the relationships between similarity maxima obtained from different molecular fields, provides another perspective of molecular matching studies in drug design application.

Application to Ligand Binding Problem

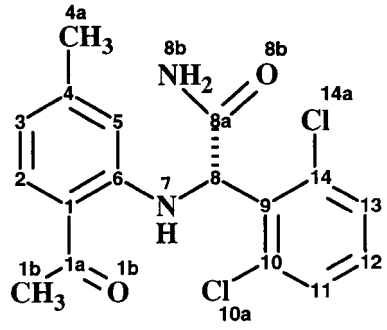
All of the described features of the present version of MIMIC were applied to a study of the



SCHEME 1. Adapting molecule rotation schematic.



Nevirapine



α-APA

SCHEME 2. Reverse transcriptase inhibitors Neurapine and α-APA.

relative ligand-binding orientation between two NNRTIs,[61] Nevirapine[62] and α-APA.[63] Scheme 2 depicts these inhibitors, which can be taken as representatives of the larger, structurally diverse class of molecules in the NNRTI family. A molecular-field matching study of a larger set of this class of inhibitors will be presented elsewhere in the due course.[64]

A relative binding orientation between Nevirapine and α-APA was recently obtained by superpositioning the surrounding protein crystal structure in each complex.[65] Thus, this proposed ligand overlay, which is based on experimentally obtained high resolution structures of HIV-1 reverse transcriptase with these inhibitors, appears to provide an interesting test of the type of results that can be achieved from the molecular-matching methodology as implemented in MIMIC. The step by step procedure used in the current molecular similarity study is presented in the following sections.

CONFORMATIONAL ANALYSIS AND CLUSTERING

The structural rigidity of Nevirapine made it an appropriate choice as the reference molecule. Therefore, α-APA was taken as the adapting molecule. Due to α-APA's structural flexibility, the first compulsory step in the similarity study was a conformational analysis of α-APA. The MOSAIC program[66] was used to explore the conformational space of α-APA by systematically ro-

tating all flexible bonds, followed by energy minimization using the MM2* molecular mechanics force field,[67,68] as implemented in the original MACROMODEL program.[66] Mulliken atomic charges for each of the lowest energy conformations were obtained using MOPAC at the AM1 semiempirical level.[69]

After performing a systematic exploration of the α-APA's conformational space, a restricted set of low-energy molecular conformations was selected. In the present case, the first 20 conformations were taken, covering an energy range of about 5 kcal/mol from the lowest energy conformation. However, within this 20-conformation set, different subsets of conformers may belong to similar regions of conformational space. Thus, to identify structural commonalities among these conformations, a *conformational similarity analysis*, based upon steric-similarity indices, was carried out by pairwise matching all 20 conformations. The similarity values obtained were then collected into a conformational similarity matrix. After appropriately transforming this similarity matrix, a principle component analysis was performed. The resultant clustering is depicted in Figure 1 using the first three principle components, which account for 87.7% of the total variance. Five different conformational clusters are clearly recognizable and are labeled according to their proximity to the conformational space covered by each cluster: cluster C1 contains 7 conformations and includes the two lowest energy conformations; cluster C2′ contains 3 conformations and is found in close proximity to cluster C2″, which collects 4 conformations; finally, clusters C3′ and C3″ contain 3 conformations each and appear to also be closely related (because their relative positions mainly differ along the PC3 axis).

From this conformational clustering, the lowest energy conformation in each cluster was taken as the representative structure of the conformational space spanned by the set of conformations within the cluster. Thus, five prototype conformations (1, 3, 4, 6, and 15, according to the energy ordering) were selected, and their 3-dimensional structures are shown in Figure 2.

Values of the characteristic dihedral angles corresponding to each of the prototype conformations are gathered in Table I. Basically, a particular

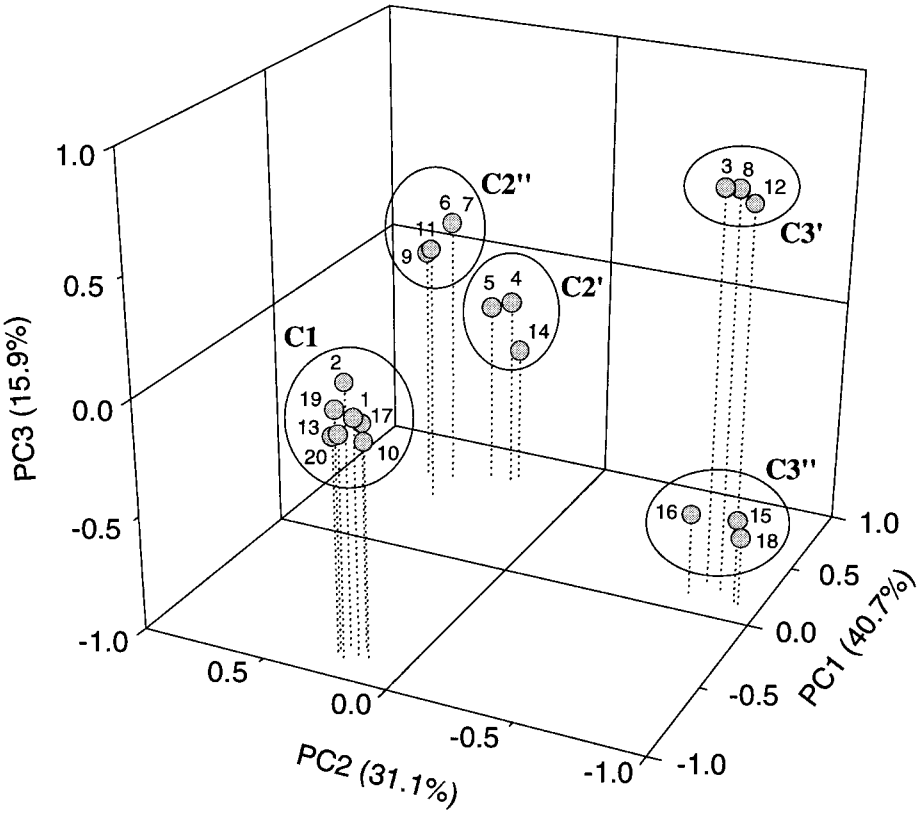


FIGURE 1. Principle component analysis of α-APA conformations.

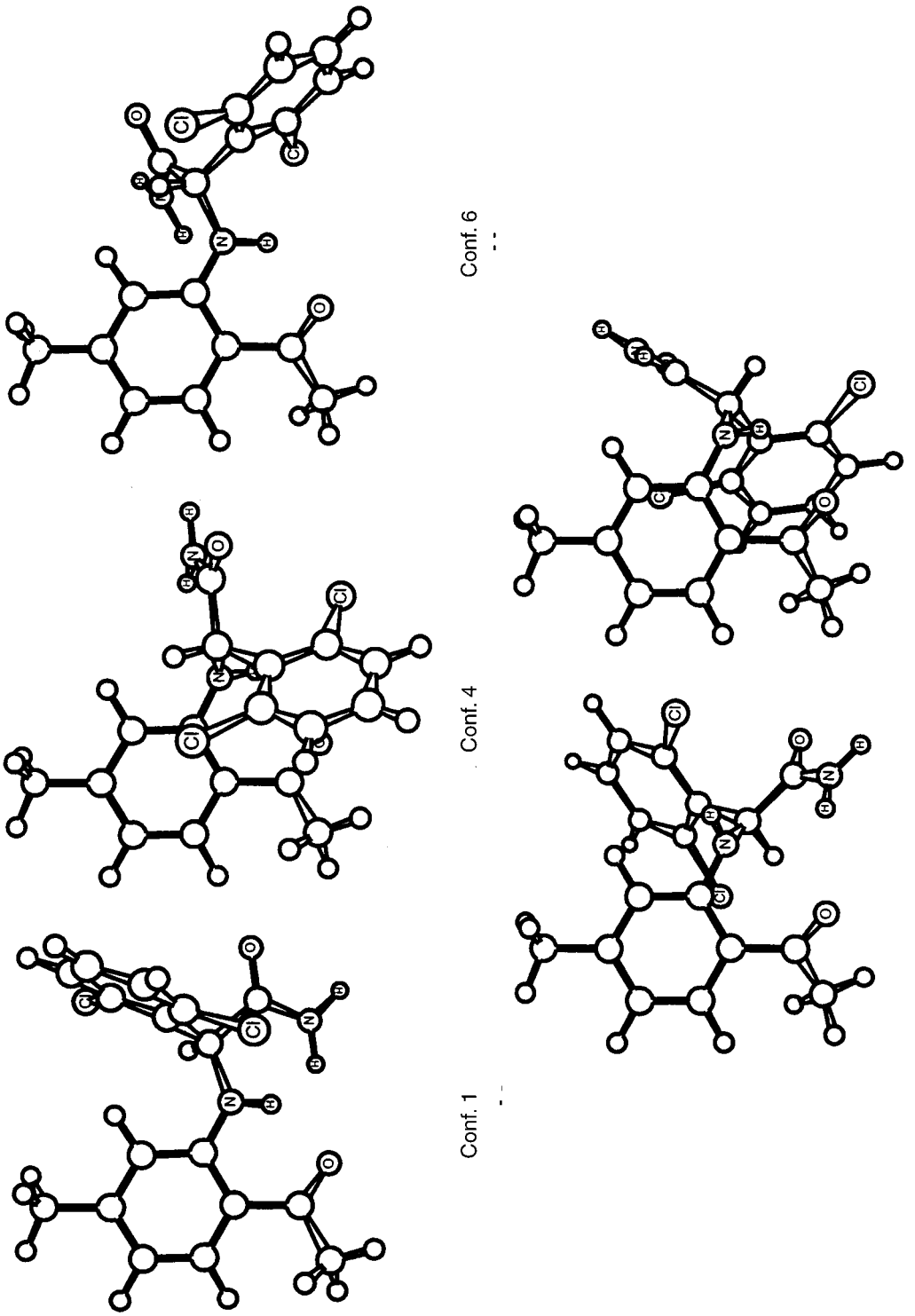


FIGURE 2. Structures of the selected five prototype α-APA conformations.

TABLE I.
Selected Dihedral Angles for Five Prototype Conformations of α-APA.

Dihedral Angle (°)	Conf. 1	Conf. 3	Conf. 4	Conf. 6	Conf. 15
C1-C6-N7-C8	153.5	72.5	−103.0	−167.1	97.9
N7-C8-C9-C14	−126.4	−107.9	−100.6	−110.9	−93.3
C6-N7-C8-C9	84.4	70.8	87.4	154.5	−60.8
C6-N7-C8-C8a	−148.5	−159.1	−142.7	−78.3	78.2
N7-C8-C8a-O8b	−153.5	−153.3	−166.5	178.9	−155.6
C6-C1-C1a-O1b	−25.7	29.2	−29.7	28.4	31.2

See Scheme 2 for atom labels.

structural feature differentiates conformations 1, 4, and 6 from conformations 3 and 15: in conformations 1, 4, and 6 the dichlorine-benzene ring is found above the plane of the paper, while for conformations 3 and 15 the same ring is located below the plane of the paper (Fig. 2). Figure 2 also allows one to visually recognize the proximity between the conformational regions being represented by each of the prototype conformations.

EXPLORATION OF SIMILARITY SPACE

A series of spherical systematic similarity searches were performed with Nevirapine as the reference molecule and each one of the α-APA prototype conformations (Fig. 2) as the adapting molecule. (In all cases, a molecular overlay was considered optimized when the gradient norm of the similarity index was below 0.0001.) A summary of the values for the highest similarity index maxima found (thus, corresponding to the best molecular overlay), using different weightings for the steric and the electrostatic contributions to the molecular similarity (denoted as steric:electrostatic), is presented in Table II. Also collected in Table II are the number of distinct similarity maxima obtained using a 45° rotational step during the spherical systematic search.

Some interesting issues become evident when analyzing the results of Table II for specific steric: electrostatic weightings of the molecular similarity. First, observe that the relative order of the similarity indices obtained for the five α-APA conformations when superimposed against Nevirapine does not necessarily conform with the order of the α-APA conformational energies. This is not surprising because there is, in principle, no reason for such a relationship to exist. Second, the highest similarity between Nevirapine and α-APA was achieved from different α-APA conformations, depending in the coefficients of the steric and electrostatic molecular fields contributing to the similarity [eq. (11)]. These two points reveal the need, when performing molecular-field matchings, to cover a wide range of the conformational space spanned by the molecules. As derived from the results in Table II, this requirement can be accomplished by use of a set of structures representative of the energetically most important conformational regions of the molecules (Fig. 1).

TABLE II.
Values of Best Similarity-Index Maxima Between Nevirapine and Five Prototype Conformations of α-APA Using Different Molecular Similarities.

Conformation	ΔE[a]	1.00:0.00[b]	0.75:0.25[b]	0.50:0.50[b]	0.25:0.75[b]	0.00:1.00[b]
1	0.00	0.6380 (51)	0.6341 (41)	0.6334 (35)	0.6422 (21)	0.6767 (14)
3	0.82	0.6094 (67)	0.6019 (61)	0.5959 (44)	0.5981 (21)	0.6312 (15)
4	0.97	0.6193 (56)	0.6089 (51)	0.6330 (43)	0.6964 (20)	0.7638 (13)
6	1.51	0.5885 (54)	0.5881 (46)	0.5996 (35)	0.6219 (27)	0.6853 (20)
15	3.64	0.5978 (64)	0.5908 (55)	0.5891 (41)	0.6443 (30)	0.7052 (19)

[a] Energy differences are given in kcal / mol.
[b] Weighting parameters of the combined molecular similarity (steric:electrostatic).

Inclusion of conformational flexibility in the matching procedure can extend or enlarge the conformational region represented by each prototype conformation and eventually may permit migration to other nearby conformational regions (see Fig. 1). This may reduce the need for extension of the conformational analysis prior to molecular similarity analysis and the number of prototype conformations that must be considered. Analysis of the role of conformational flexibility is currently being explored in our laboratory.

The validity of grid-based cumulative similarity indices [eq. (10)] was tested by comparison to values obtained from integration, over all space, of the corresponding molecular fields; the results are summarized in Table III. Two different grid spacings, 0.5 and 1.0 Å, are employed for the sake of comparison. As deduced from the table, steric-similarity indices are accurately reproduced with the 0.5Å grid-based similarity indices, while a 1.0-Å grid spacing did not even produce qualitatively reasonable results for these indices. On the other hand, the quality of grid-based electrostatic-similarity indices was similar for both grid spacings, which reproduced the same ordering obtained in Table II. These results strongly recommended that a 0.5Å grid spacing be used in calculations of grid-based cumulative steric-similarity indices, while the coarser 1.0Å grid seemed to be sufficient for obtaining qualitatively good results for grid-based cumulative electrostatic-similarity indices.

Examination of the similarity indices presented in Table II for the five different molecular field weighting combinations scanned, allows ranking of the α-APA conformations based on their ability to match the steric or electrostatic field of Nevirapine. The results show that conformation 1 of α-APA (α-APA$_1$) has the largest similarity-index value when the steric contribution to the similarity dominates over the electrostatic contribution, while conformation 4 of α-APA (α-APA$_4$) exhibits the largest similarity-index value when the electrostatic contribution to the similarity dominates over the steric contribution. When steric and electrostatic similarities were weighted equally, comparable similarity-index values were obtained for both conformations. Because of the apparent relationship of specific similarity components to particular conformation of the adapting molecule (α-APA in this case), α-APA$_1$ and α-APA$_4$ were selected for more detailed similarity analyses.

Molecular overlays corresponding to the largest steric-similarity index between Nevirapine and α-APA$_1$ (0.6380) and the largest electrostatic-similarity index between Nevirapine and α-APA$_4$ (0.7638) are depicted in Figure 3. The most relevant steric and electrostatic features common to both molecules are clearly manifested. For the sake of simplicity, the following notation, [Nevirapine atom or group|α-APA atom or group], will be used to indicate close atomic ''contacts'' between the two molecules. From the best steric-similarity overlay (Fig. 3a) several common structural patterns are evident: [O13|Cl14a], [N10|Cl10a], [N11|N7], and [4-methyl pyridine ring|1-methoxy-4-methyl benzene ring] (see Scheme 2 for atomic labeling). On the other hand, the best electrostatic-similarity overlay (Fig. 3b) shows that the two molecules appear to be reorganized according to their common electrostatic features. Different atomic correspondences are now evident; [O13|O8b], [N5|N7 (with N8b also pointing toward this electrostatic region)], [cyclopropane ring|O1b], and the two six-member rings in close proximity to each other. Thus, it is clear from Figure 3 that using different

TABLE III.
Single-Point Grid-Based Cumulative Similarity Indices at Final Relative Orientations Found from Molecular Matchings Using Continuum Similarity Indices.

Conformation	1.00:0.00[a]	0.50:0.50[a]	0.00:1.00[a]
1	0.6380 (0.6246)	0.6359 (0.6253)	0.6756 (0.6756)
3	0.6094 (0.6156)	0.6094 (0.6098)	0.6192 (0.6193)
4	0.6193 (0.6178)	0.6297 (0.6268)	0.7571 (0.7569)
6	0.5885 (0.5867)	0.5955 (0.6026)	0.6931 (0.6925)
15	0.5978 (0.5821)	0.5914 (0.5831)	0.7073 (0.7064)

Indexes found in (Table II). Calculations were done using a 0.5-Å grid spacing (ca. 24,000 grid points). Results obtained using a 1.0-Å grid spacing (ca. 3000 points) are also added in parentheses for the sake of comparison.
[a] Weighting parameters of the combined molecular similarity (steric:electrostatic).

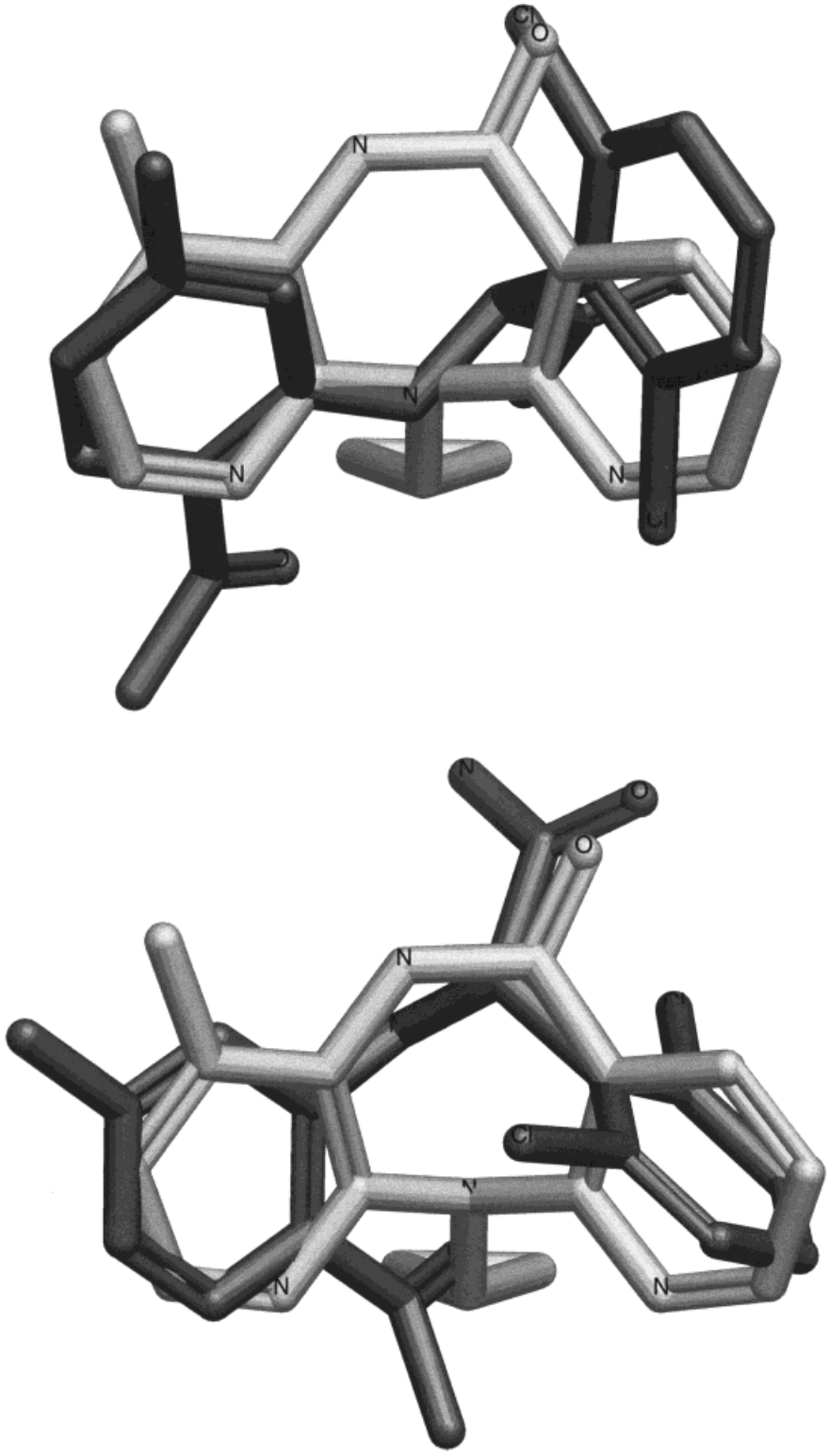


FIGURE 3. (a) Best steric similarity molecular overlay between Nevirapine and conformation 1 of α-APA and (b) best electrostatic similarity molecular overlay between Nevirapine and conformation 4 of α-APA.

molecular fields to compute the similarity between molecules gives rise to different molecular overlays, which may expose essentially novel common features.

Comparison of the experimentally derived overlay between Nevirapine and α-APA[65] with the two theoretically derived overlays depicted in Figure 3 reveals that the steric-based overlay (Fig. 3a) is in closest agreement with the experimental overlay. Thus, at least for this particular case and under the molecular-field approximations used, the steric contribution is more important than the electrostatic contribution in reproducing the correct relative ligand-binding orientation between Nevirapine and α-APA. This is a result that should be taken into consideration when extending the molecular matching study to other members of the NNRTI family.[64]

SIMILARITY SPACE COVERAGE OF SIMILARITY MAXIMUM

The set of molecular overlays obtained from a given molecular field is not likely to be the same as that obtained using a different molecular field (Table II). In this sense, each type of molecular field based similarity induces its own *molecular similarity space*, within which its characteristic similarity hypersurface is embedded. As similarity optimizations seek to maximize molecular similarity-index values, surrounding similarity maxima regions correspond to ''hills'' rather than basins, such as in energy minimizations on potential energy hypersurfaces.[70]

An interesting result that arises from the data given in Table II is that molecular similarity spaces dominated by the steric-field component exhibit many more local similarity maxima than those dominated by the electrostatic-field component. For example, in the case where a pure steric field is used, $\alpha\text{-APA}_1$ and $\alpha\text{-APA}_3$ exhibit 51 and 67 local similarity maxima, respectively. In contrast, when a pure electrostatic field is used, the same two α-APA conformations yield only 14 and 15 local similarity maxima, respectively. This is not entirely surprising due to the fact that molecular similarity derived from steric fields depends primarily upon relatively local (short-range) overlaps among atomic-centered steric-field contributions. On the other hand, molecular similarity derived from electrostatic fields depends upon short- and long-range contributions, which define wider regions of space about a molecule in terms of reactivity properties and thus substantially reduce the number of local similarity maxima. However, even when the number of local similarity maxima is quite large in cases where the steric-field contribution to the similarity dominates, the soft-Gaussian approach employed in the MSV representation used here [see eqs. (1), and (2)] appears to provide a relatively efficient means for avoiding one of the major problems of electron density based molecular similarity, becoming trapped in structurally irrelevant molecular overlays.[50]

Another point worth exploring is the nature of the regions surrounding similarity maxima. Each such region is separated from other regions by hyperdimensional valleys. Because gradient-based procedures are used to locate similarity maxima, an optimization initiated at any point lying within the hill region corresponding to a given similarity maximum will terminate at that maximum. Then, regions corresponding to similarity maxima can be explored using the results of systematic searches based upon a large set of uniformly generated starting orientations.

The number of times that a particular similarity maximum is obtained during a systematic search is a measure, albeit an approximate one, of the importance and relative isolation of a given similarity maximum. The greater the number of ''hits,'' the larger the ''hill region'' corresponding to that maximum. Figure 4 depicts the frequency histograms derived from the 51 steric-similarity maxima (1.00:0.00) obtained for $\alpha\text{-APA}_1$ (Fig. 4a) and the 13 electrostatic-similarity maxima (0.00:1.00) located for $\alpha\text{-APA}_4$ (Fig. 4b), respectively. From the figure it is seen that the largest frequencies of occurrence (in percent) coincide with the respective best steric- and electrostatic-similarity maxima, 0.6380 and 0.7638, respectively. (Although this is true for the two particular cases presented, the same may not necessary hold in other cases.) This implies that these maxima ''cover'' larger regions of similarity space than do other maxima obtained in the similarity study. Importantly, although other similarity maxima are encountered, the frequency values corresponding to the different similarity maxima do not change appreciably when a rotational step size smaller than the 45° step used to generate the histograms in Figure 4 is employed. This is especially true when the electrostatic field contributes preferentially to the similarity, for the reasons described above. As an example, tighter step sizes of 30° (744 starting

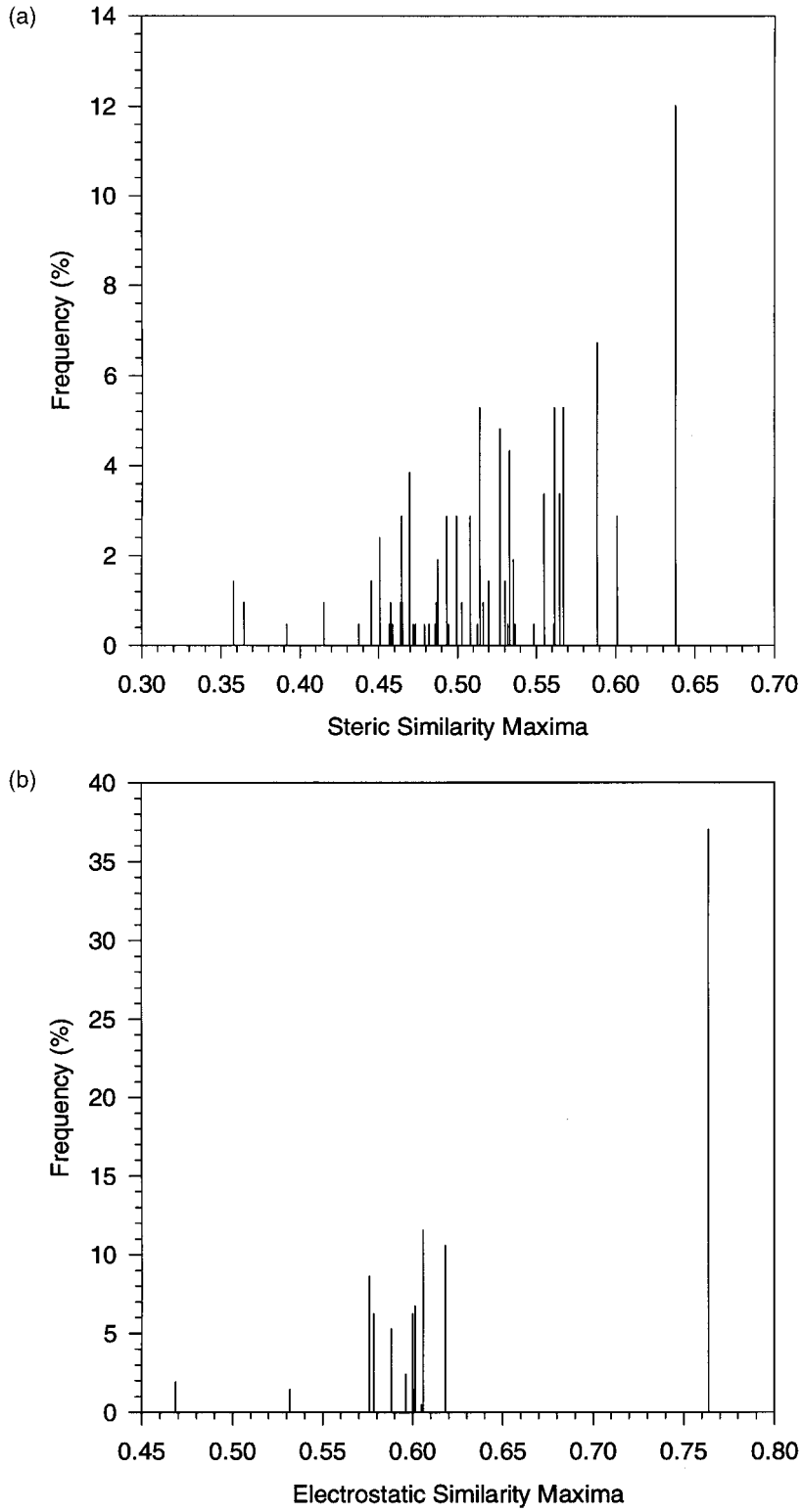
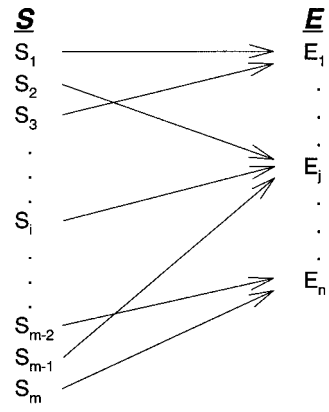


FIGURE 4. Frequency histograms from the spherical systematic seraches for (a) the steric similarity maxima between Nevirapine and conformation 1 of α-APA and for (b) the electrrostatic similarity maxima between Nevirapine and conformation 4 of α-APA.

orientations), 20° (2916 starting orientations), and 10° (22,104 starting orientations) were selected to explore the steric-similarity space between Nevirapine and α-APA$_1$. A comparison of these results with those obtained from the systemic search using the 45° step size (208 starting orientations) shows that, although the number of steric similarity maxima continued to increase gradually (56, 66, and 79 for 30°, 20°, and 10°, respectively), most of the additional maxima found had low similarity index values and very low reaching frequency, which did not change the essential results shown in Figure 4a. This clearly confirms the robustness of the procedure.



SCHEME 3. Relationship between two similarity spaces defined by two different sets of molecular overlays.

RELATIONSHIPS AMONG MOLECULAR-FIELD SIMILARITY SPACES

The study of the relationships among similarity spaces induced by different types of molecular fields are also of interest and have yet to be exploited in any significant way. Scheme 3 illustrates the kind of relationship that may be found between two similarity spaces defined by two different sets of molecular overlays. Here attention is focused on the relationship between the steric- and electrostatic-similarity spaces, denoted as $\mathbf{S} = \{S_i\}$ and $\mathbf{E} = \{E_j\}$, respectively, in the scheme. The set of $\{S_i\}$ indicated on the left-hand side of the diagram corresponds to the steric-similarity molecular overlays obtained by the exploration of the steric-similarity space. Each of these S_i steric-similarity overlays is then taken as a starting orientation for reoptimization based solely upon electrostatic similarity to yield a given similarity maximum E_j belonging to the set of $\{E_j\}$ electrostatic-similarity molecular overlays. This reoptimization process is indicated by an arrow having its origin in a S_i and pointing toward the connected E_j. As is evident from the data in Table II, due to the fact that generally steric-similarity spaces possess a larger number of similarity maxima than electrostatic-similarity spaces (thus in Scheme 3, $m \geq n$ in general), several S_i may converge to the same E_j. Proceeding in this way, this analysis will be able to identify the set of steric-similarity overlays that belong to the hill region of a unique electrostatic-similarity overlay. This linkage between a given overlay structure in electrostatic-similarity space with several overlay structures in steric-similarity space can be exploited in the design of new strategies for determining the best molecular overlay based upon steric and electrostatic similarities.

The above-mentioned connectivity between a pair of molecular overlays belonging to different similarity spaces can be analyzed by constructing the type of diagrams presented in Figure 5. These diagrams are intended to provide a useful, visual means to better understand the relationships among similarity spaces. Figure 5a depicts the relationship between the steric- and electrostatic-similarity spaces defined by the similarity of Nevirapine with α-APA$_1$; Figure 5b depicts the corresponding results with α-APA$_4$. According to the procedure described above, every point in Figure 5 and connects each steric-similarity maximum (defined by projection of the point onto the steric-similarity space axis) with an electrostatic-similarity maximum (defined by projection of the point onto the electrostatic-similarity space axis).

Analysis of Figure 5a reveals several interesting features. The top right point in this figure corresponds to a steric-similarity index value of 0.6380, which is the best possible steric-similarity overlay between Nevirapine and α-APA$_1$. The electrostatic-field similarity optimization taking this steric-similarity overlay as the starting orientation leads to an electrostatic-similarity index value of 0.6695, which is close to the best possible electrostatic-similarity index value between Nevirapine and α-APA$_1$ (0.6767) but it is actually the second best value. On the other hand, as seen in the top row of Figure 5a, five different steric similarity maxima (with values 0.4453, 0.4576, 0.4651, 0.4698, and 0.5270) do lead to the best electrostatic-similarity index value of 0.6767. Interestingly, the largest steric-similarity maximum (0.5270) that gave rise to the best electrostatic-similarity maximum (0.6767) is ranked 15th.

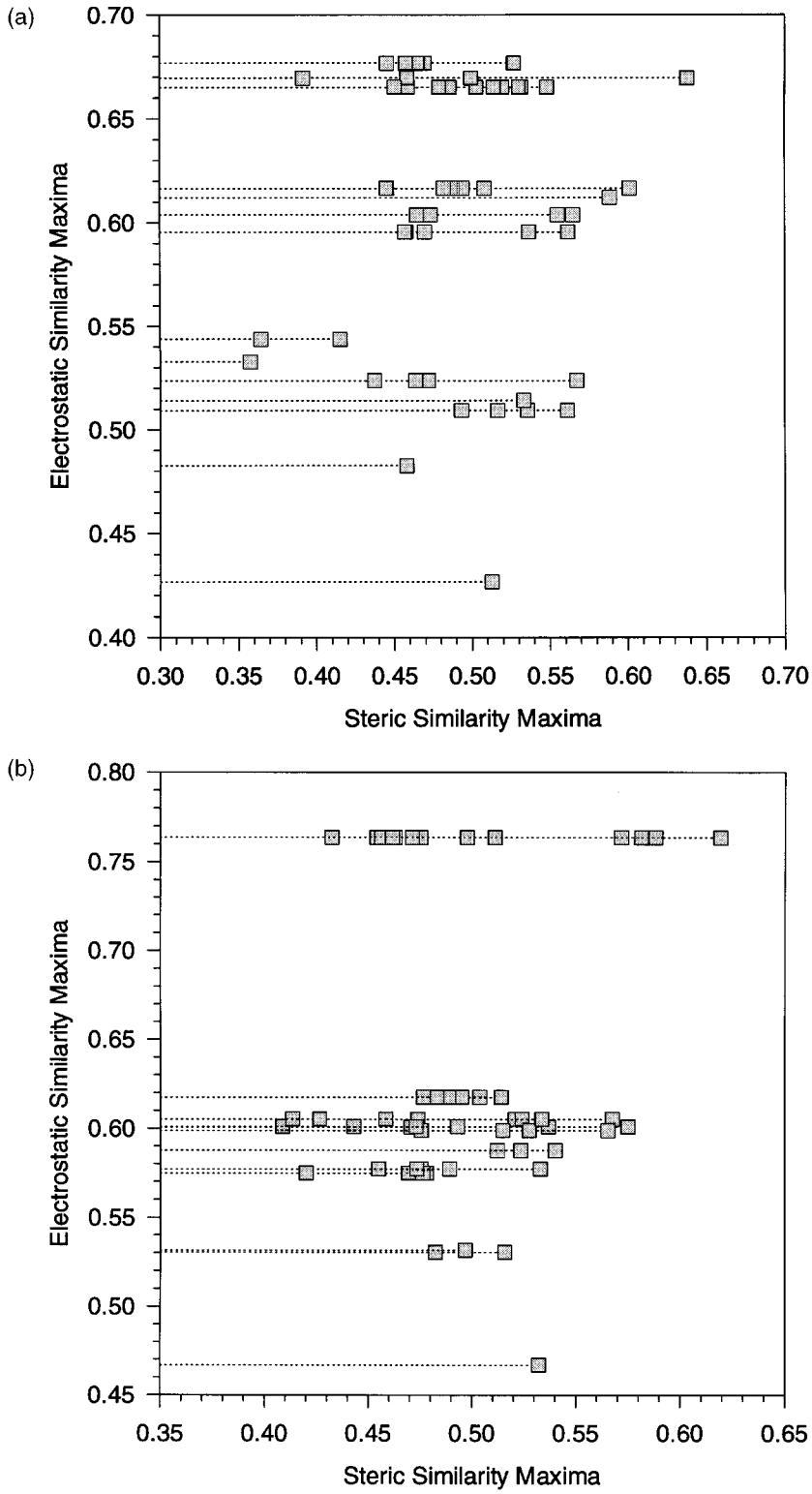


FIGURE 5. Similarity-space correlation diagrams between the steric- and the electrostatic-similarity spaces between Nevirapine and α-APA for (a) conformation 1 and (b) conformation 4 of α-APA.

Figure 5b exhibits analogous behavior to that observed in Figure 5a, except that in this case the best steric-similarity overlay of α-APA$_4$ is associated with the best electrostatic-similarity overlay (0.7638). Moreover, in this case the first three steric-similarity maxima (with values 0.6193, 0.5883, and 0.5816) also belong to the hill region of the best electrostatic-similarity overlay. Of interest is the fact that the first steric-similarity maximum not associated with the best electrostatic-similarity maximum (which is actually the fourth best steric-similarity overlay, with a value of 0.5751) is connected to the fifth best electrostatic-similarity overlay (0.6015).

To illustrate the connection between similarity maxima belonging to different molecular-field similarity spaces. Figure 6 depicts the different overlays between Nevirapine and α-APA$_1$ that were obtained by systematically increasing the electrostatic contribution to the similarity of the molecular overlay corresponding to the best steric-similarity overlay (0.6380). In this sense, it is possible to visually follow the changing orientation of the molecular overlay connecting the best steric-similarity maximum and its related electrostatic-similarity maximum (which corresponds to the steric-electrostatic overlay connection defined by the top right point in Fig. 5a). As can be observed, introducing an electrostatic contribution to the steric-similarity overlay does not disrupt the essence of the molecular overlay, but it helps to modulate its relative orientation.

A more detailed analysis of Figure 5 clearly shows that the electrostatic-similarity maxima can also be grouped into different electrostatic-similarity clusters, which appear as ''levels'' in this type of diagrams. For example, the top three electrostatic-similarity maxima (with values 0.6767, 0.6695, and 0.6650) constitute a neatly differentiated level from the rest of the electrostatic-similarity maxima. Interestingly, visual examination of all overlays belonging to this level (steric and electrostatic) revealed a common structural feature, namely that the dichlorine-benzene ring of α-APA$_1$ is always located at the pyridine side of Nevirapine (as depicted in Fig. 6). However, the molecular overlays constituting the second level of overlays had the same dichlorine-benzene ring of α-APA$_1$

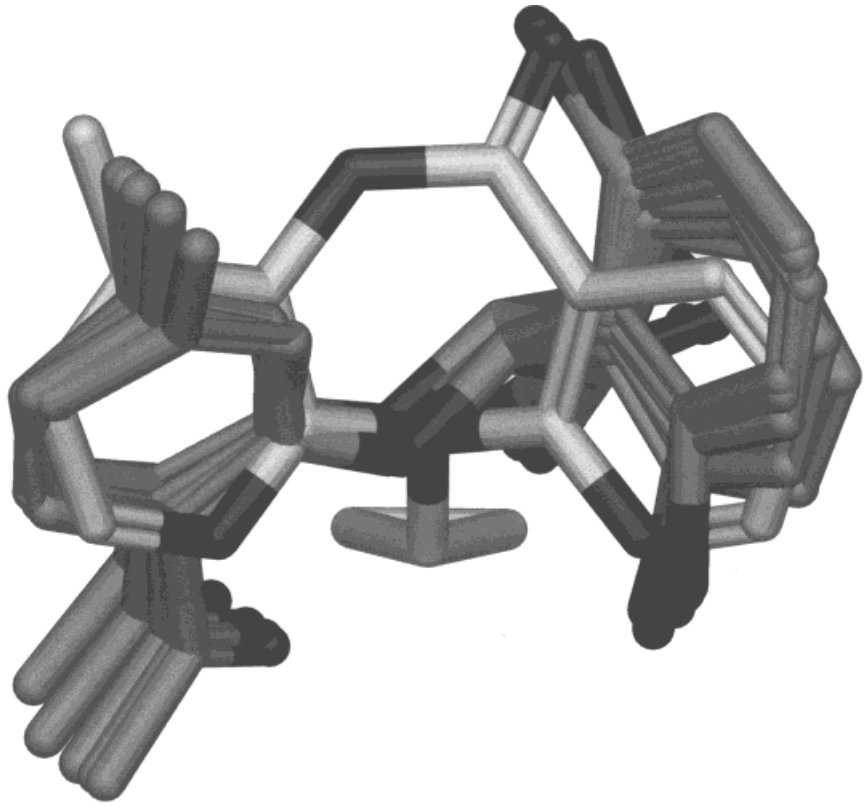


FIGURE 6. Modulation of the molecular overlay between Nevirapine and conformation 1 of α-APA due to a systematic introduction of an electrostatic contribution into the similarity index.

located at the 4-methyl pyridine side of Nevirapine. This interesting finding shows that these kinds of diagrams may also serve to expose (in a wider sense) the possibility of different preferential modes of overlay between Nevirapine and α-APA.

LOCATION OF MAXIMUM SIMILARITY LOCI

As qualitatively described above, common structural patterns between Nevirapine and α-APA can be recognized visually from the best steric-similarity matching depicted in Figure 3a. However, calculation of the atomic contributions to the total molecular similarity [eq. (7)] can be used to quantitatively ellucidate those sites of each molecule where the steric similarity is concentrated. The atomic contributions for each molecule (given in percentage of the total steric similarity, 0.6380) are presented in Figure 7. The atoms contributing most to the total steric similarity for Nevirapine are O13 (14.7%), N10 (9.5%), and N11 (8.2%), yielding 32.4% of Nevirapine's contribution to the overall steric similarity. The most important contributions in α-APA, Cl14a (15.5%), Cl10a (13.4%), and N7 (8.8%), give rise to 37.7% of α-APA's contribution to the total steric similarity. Thus, only three atoms in each molecule contribute about one-third of the total steric similarity between the two molecules. These atoms constitute what can be called the *structural anchors* of each molecule as these atomic loci strongly determine the final 3-dimensional orientation of the best steric overlays, namely, those having large steric-similarity index values. Another major contribution to the total steric similarity is the overlay of the Nevirapine's 4-methyl pyridine ring and α-APA's 1-

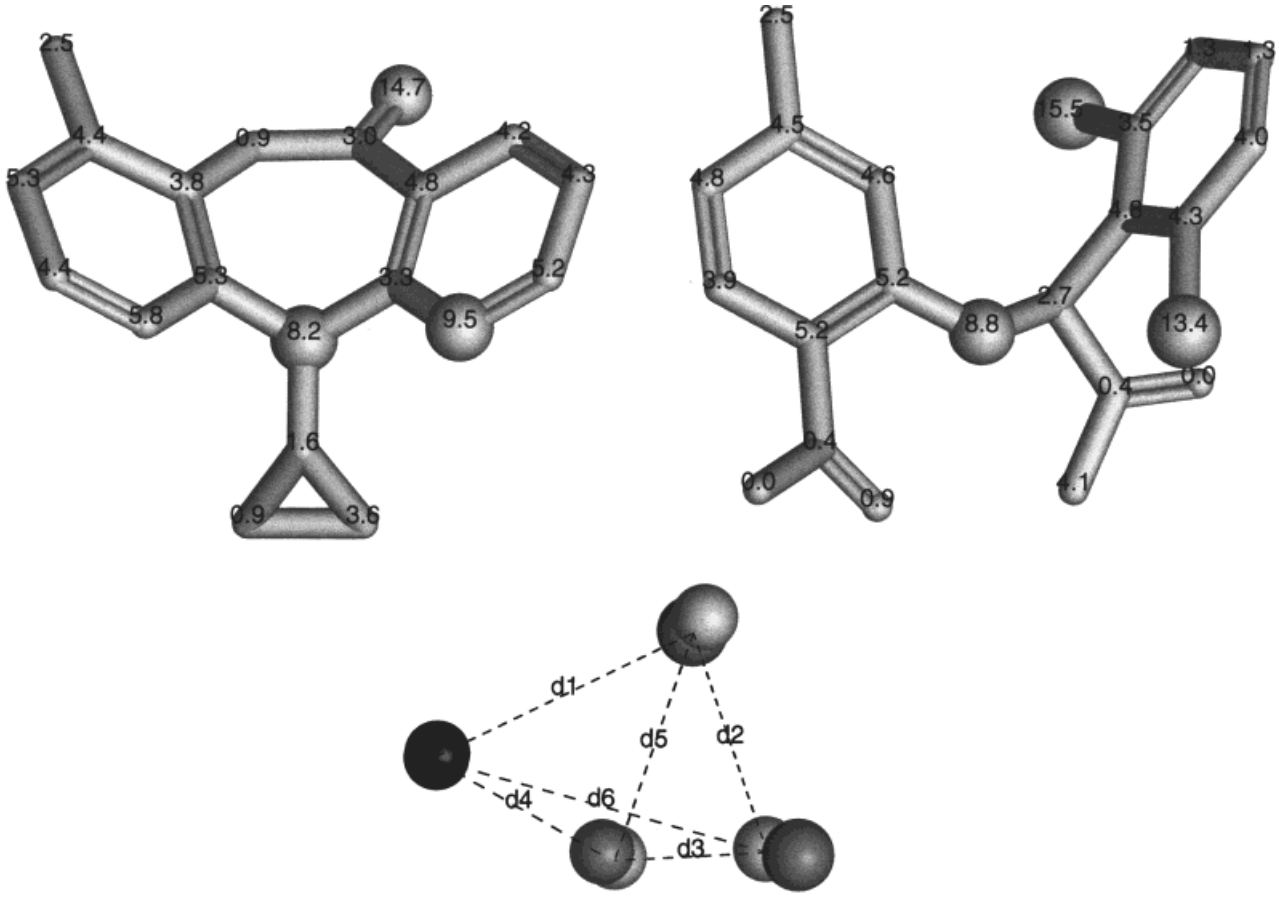


FIGURE 7. Atomic contributions (%) to the total steric similarity index value between Nevirapine and conformation 1 of α-APA. The structual pharmacophore pattern collects the structural anchors of the two molecules (lighter and darker atoms belong to Nevirapine and α-APA, respectively). The black balls indicate the centers of the two left-hand six-membered rings.

methoxy-4-methyl benzene ring (see Fig. 7), which contribute an additional 29.0% to the steric similarity in Nevirapine and 28.2% in α-APA.

Location of these *maximum similarity loci* may be of great help when attempting to construct a common structural pharmacophore pattern among a set of molecules. Under this discrete approach, construction of a common pattern between Nevirapine and α-APA can be accomplished by linking the three structural anchors and the common aromatic rings, as shown schematically in Figure 7. Six distances were selected to connect each maximum similarity to loci. The average values and tolerance of d1, d2, d3, d4, d5, and d6 are 4.98 ± 0.01, 5.14 ± 0.37, 2.75 ± 0.30, 2.72 ± 0.10, 4.18 ± 0.10, and 4.89 ± 0.25 Å, respectively. These distances define the 3-dimensional relationships among the atomic loci where the steric similarity is essentially concentrated and constitute the bases for schematically representing the main common steric patterns between the two molecules. The present model represents only a simple example of the first step in the construction of a common structural pharmacophore pattern for a set of ligands that all bind to the same receptor site. Improvement of the model can be achieved systematically through the inclusion of additional structural features deduced by consideration of other ligands in the similarity study. Thus, the model itself should not be regarded as unique and static.

Conclusions

The development of MIMIC was presented, including a description of the basic methodological concepts and their algorithmic implementations that allow the program to perform similarity matchings and analyses quickly and accurately. In addition, several original molecular similarity approaches for conformational clustering, exploration of similarity spaces together with the analysis of their overlay solutions and interconnections, and location of maximum similarity loci were proposed that permit a deeper examination of the overall similarity. Finally, new methodological features and similarity procedures were suggested by results obtained from these approaches, which are currently being implemented in MIMIC and will be the subject of future discussions.

The relative ligand-binding orientation between Nevirapine and α-APA was selected as a case study to demonstrate the performance of MIMIC and the molecular similarity approaches employed. The best molecular overlays obtained in the similarity study were compared against those reported from X-ray crystallographic data. These comparisons showed that the experimental overlay was found within the best set of overlays determined by the molecular field based similarity methodology described here. In many cases, however, experimentally derived overlays are not available, and it is necessary to be able to determine the correct overlay from the set of computed overlays. This is especially difficult in situations where the molecules being overlayed are structurally diverse, as is the case for the NNRTIs investigated here. This Nevirapine and α-APA study clearly showed the complexity of deriving a relative ligand-binding overlay model, which involves analysis of multiple conformations for each molecule and multiple overlays for each conformation. The application of molecular-similarity approaches to the multiple conformation problem allowed selection of a small number of diverse conformations that provided a representative sample of the conformational space. The quantitative nature of the exploration of similarity spaces based on molecular-field matching then allowed ranking of the molecular overlay results obtained from the selected conformations. This provided a means of selecting the set of matches that contained the correct overlay. While the complexity of the study increased with the number of different molecular-fields employed, the diversity of results provided insight into the nature of the breadth of the matches obtained. For these reasons, a maximum exploitation of the similarity results from several points of view was necessary and provided useful clues to the best relative orientation or best set of molecular overlays. Even in the latter case, structure-based drug design projects can benefit from this information as a guide for improving the lead compounds, making molecular field based similarity an important approach supporting the drug design.

Acknowledgments

We would like to thank Dr. James D. Petke for many helpful comments and discussions. J. M. acknowledges the financial support provided by

the Generalitat de Catalunya through a CIRIT grant (No. 1995BEAI200115).

References

1. M. A. Johnson and G. M. Maggiora, Eds., *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990.
2. H. Kubinyi, Ed., *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993.
3. K. Sen, Ed., *Topics in Current Chemistry*: *Molecular Similarity I*, vol. 173, Springer−Verlag, Berlin, 1995.
4. P. M. Dean, Ed., *Molecular Similarity in Drug Design*, Blackie Academic, London, 1995.
5. R. Carbó, Ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer, Amsterdam, 1995.
6. A. C. Good and W. G. Richards, *J. Chem. Inf. Comput. Sci.*, **33**, 112 (1993).
7. B. B. Masek, A. Merchant, and J. B. Matthew, *Proteins: Struc. Function Genet.*, **17**, 193 (1993).
8. J. S. Tokarski and A. J. Hopfinger, *J. Med. Chem.*, **37**, 3639 (1994).
9. P. G. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH, New York, 1993.
10. P. Politzer and D. G. Truhlar, Eds., *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
11. R. P. Apaya, B. Lucchese, S. L. Price, and J. G. Vinter, *J. Comput.-Aided Mol. Design*, **8**, 653 (1994).
12. J. G. Vinter and K. I. Trollope, *J. Comput.-Aided Mol. Design*, **9**, 297 (1995).
13. C. Burt, W. G. Richards, and P. Huxley, *J. Comput. Chem.*, **11**, 1139 (1990).
14. A. M. Richard, *J. Comput. Chem.*, **12**, 959 (1991).
15. F. Manaut, F. Sanz, J. Jose, and M. Milesi, *J. Comput.-Aided Mol. Design*, **5**, 371 (1991).
16. A. C. Good, E. E. Hodgkin, and W. G. Richards, *J. Chem. Inf. Comput. Sci.*, **32**, 188 (1992).
17. J. D. Petke, *J. Comput. Chem.*, **14**, 928 (1993).
18. A. K. Ghose and G. M. Crippen, *J. Comput. Chem.*, **9**, 80 (1988).
19. P. Furet, A. Sele, and N. C. Cohen, *J. Mol. Graph.*, **6**, 182 (1988).
20. V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, *J. Chem. Inf. Comput. Sci.*, **29**, 163 (1989).
21. P. Croizet, M. H. Langlois, J. P. Dubost, P. Braquet, E. Audry, Ph. Dallet, and J. C. Colleter, *J. Mol. Graph.*, **8**, 153 (1990).
22. A. Kantola, H. O. Villar, and G. H. Loew, *J. Comput. Chem.*, **12**, 681 (1991).
23. W. Heiden, G. Moeckel, and J. Brickmann, *J. Comput.-Aided Mol. Design*, **7**, 503 (1993).
24. P. Gaillard, P.-A. Carrupt, B. Testa, and A. Boudon, *J. Comput.-Aided Mol. Design*, **8**, 83 (1994).
25. I. Rozas, Q. Du, and G. A. Arteca, *J. Mol. Graph.*, **13**, 98 (1995).
26. N. Kurochkina and B. Lee, *Protein Eng.*, **8**, 437 (1995).
27. R. G. A. Bone and H. O. Villar, *J. Mol. Graph.*, **13**, 201 (1995).
28. P. M. Dean and P.-L. Chau, *J. Mol. Graph.*, **5**, 152 (1987).
29. Y. Kato, A. Itai, and Y. Iitaka, *Tetrahedron*, **43**, 5229 (1987).
30. R. Cramer III, D. Patterson, and J. Bunce, *J. Am. Chem. Soc.*, **110**, 5959 (1988).
31. R. Carbó and B. Calabuig, *Comput. Phys. Commun.*, **55**, 117 (1989).
32. S. K. Kearsley and G. M. Smith, *Tetrahedron Comput. Methods*, **3**, 615 (1990).
33. R. B. Hermann and D. K. Herron, *J. Comput.-Aided Mol. Design*, **5**, 511 (1991).
34. F. Sanz, F. Manaut, J. Rodríguez, E. Lozoya, and E. López-de-Briñas, *J. Comput.-Aided Mol. Design*, **7**, 337 (1993).
35. J. Mestres, M. Solà, M. Duran, and R. Carbó, *J. Comput. Chem.*, **15**, 1113 (1994).
36. P. Baricic and M. Mackov, *J. Mol. Graph.*, **13**, 184 (1995).
37. D. C. Rohrer, In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. Carbó, Ed., Kluwer, Amsterdam, 1995, p. 141.
38. ASP, computer software available from Oxford Molecular Ltd., Oxford, England.
39. APEX-3D, computer software available from Molecules Simulations Inc., San Diego, CA.
40. SPARTAN, computer software available from Wavefunction Inc., Irvine, CA.
41. C. McMartin and R. S. Bohacek, *J. Comput.-Aided Mol. Design*, **9**, 237 (1995).
42. T. D. J. Perkins and P. M. Dean, *J. Comput.-Aided Mol. Design* **7**, 155 (1993).
43. G. Klebe, T. Mietzner, and F. Weber, *J. Comput.-Aided Mol. Design*, **8**, 751 (1994).
44. T. I. Oprea, C. L. Waller, and G. R. Marshall, *J. Med. Chem.*, **37**, 2206 (1994).
45. A. J. Hopfinger, B. J. Burke, and W. J. Dunn III, *J. Med. Chem.*, **37**, 3768 (1994).
46. J. A. Grant and B. T. Pickup, J. Phys. Chem., **99**, 3503 (1995).
47. J. P. Ritchie and A. S. Coopenhaver, *J. Comput. Chem.*, **16**, 777 (1995).
48. R. Carbó, L. Leyda, and M. Arnau, *Int. J. Quantum Chem.*, **17**, 1185 (1980).
49. E. Besalú, R. Carbó, J. Mestres, M. Solà, *Top. Curr. Chem.*, **173**, 31 (1995).
50. J. Mestres, M. Solà, M. Duran, and R. Carbó, In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. Carbó, Ed., Kluwer, Amsterdam, 1995, p. 89.
51. G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler, and D. A. Dunn, In *Computer-Assisted Drug Design: ACS Symposium Series* 112, E. C. Olson and R. E. Christofferson, Eds., American Chemical Society, Washington, D.C., 1979, p. 205.
52. R. Carbó and E. Besalú, *Comput. Chem.*, **18**, 117 (1994).
53. S. F. Boys, *Proc. R. Soc. London*, **A200**, 542 (1950).
54. E. E. Hodgkin and W. G. Richards, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **14**, 105 (1987).

55. R. Carbó, E. Besalú, B. Calabuig, and L. Vera, *Adv. Quantum Chem.*, **25**, 253 (1994).
56. A. C. Good, *J. Mol. Graph.*, **10**, 144 (1992).
57. G. Klebe, *Perspect. Drug Discovery Design*, **3**, 85 (1995).
58. P. Willett, *Similarity and Clustering in Chemical Information Systems*, Research Studies Press, Letchworth, U.K., 1987.
59. P. S. Shenkin and D. Q. McDonald, *J. Comput. Chem.*, **15**, 899 (1995).
60. P. M. Dean and P. Callow, *J. Mol. Graph.*, **5**, 159 (1987).
61. E. De Clercq, *J. Med. Chem.*, **38**, 2491 (1995).
62. V. J. Merluzzi, K. D. Hargrave, M. Labadia, K. Grozinger, M. Skoog, J. C. Wu, C.-K. Shih, K. Eckner, S. Hattox, J. Adams, A. S. Rosenthal, R. Faanes, R. J. Eckner, R. A. Koup, and J. L. Sullivan, *Science*, **250**, 1411 (1990).
63. R. Pauwels, K. Andries, Z. Debyser, P. Van Daele, D. Schols, A.-M. Vandamme, P. Stoffels, K. De Vreese, R. Woestenborghs, C. G. M. Janssen, J. Anne, G. Cauwenberg, J. Desmyter, J. Heykants, M. A. C. Janssen, E. De Clercq, and P. A. J. Janssen, *Proc. Natl. Acad. Sci. USA*, **90**, 1893 (1993).
64. J. Mestres, D. C. Rohrer, and G. M. Maggiora, unpublished manuscript.
65. J. Ren, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Darby, Y. Jones, D. Stuart, and D. Stammers, *Nature Struct. Biol.*, **2**, 293 (1995).
66. MOSAIC is the molecular modeling program developed by the CADD group of P & U. Originally it was based on MACROMODEL; F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, and W. C. Still, *J. Comput. Chem.*, **11**, 440 (1990).
67. N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977).
68. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989).
69. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).